A Banner Ads Searching and Counting System for Sports Videos

Chueh Wei Chang#1, Shio Wen Chen*2, Hsin Fu Chien#3
Department of Computer Science and Information Engineering National Taipei University of Technology
1, Sec. 3, Chung-hsiao E. Rd., Taipei,10608,Taiwan, R.O.C.
[1]cwchang@csie.ntut.edu.tw;[3]nm58903@cc.cust.edu.tw

Abstract- **The objective of this paper was to present an approach to design a system for searching and counting the number of banner ads appearing in sports videos. The searching proceedings of matching feature points from pre-specified sample ads use the color histogram and the SURF (Speeded-Up Robust Features) algorithm. This searching approach can robustly identify objects among scaling and partial occlusion while achieving nearly real-time performance. This system worked effectively in our experiments using real videos.**

Keywords- Banner Advertising, SURF, Feature Description, Object Recognition, Image Registration

I. INTRODUCTION

Because of the large number of TV audience, the banner ads in sports videos are becoming an attractive medium for advertising. An advertiser pays an amount based on the exposures of an ad to establish good enterprise impression. The ads exposure frequency in a TV is very important for the pricing of this advertising. For example, in a baseball game, advertisers usually set up banner ads on the fences around the ball inner and outer fields. When a ball game broadcasts on the TV, the major view in most of the time is the pitching and hitting between pitcher and batter. So, the banner ad behind the home plate increases the chance of creating impression relating to ads in other places, and also it costs more expense than others.

The frequency of exposures during a watching is the number of times that the ad is displayed to the audience during the watching time. The ads exposure frequency is calculated by the frame number per time unit that the specific ads displayed. If we count this exposure frequency manually by person, it will be very time consuming and error-prone.

Because the camera shooting is taken by many different ways from many different places in a sport TV broadcast, there are several difficulties cause the ads hard to recognize, such as the similar color area, zooming factor of cameras, lighting condition in the ball fields, the occlusion by players, and partial appear, etc.

Therefore, this banner ad searching task not only becomes the problem of finding the same color area, but also finds the point correspondence between two images. In order to develop an efficient matching system, an important issue is to extract point features to describe the banner ads.

In this paper, we propose an approach to design a system that can search and count the specific banner ads in the sports videos automatically. User only needs to choose the sample images of ads. Then, this system will start parsing the whole video and record the time and show the ads appear. Our approach is based on a color filtering and the SURF feature extraction that can process near real-time and accurate.

II. RELATED WORKS

In the early years, researchers develop many image processing methods by color, texture, and local shape for content-based image retrieval [9]. Recently, the commonly used methods of feature extraction for the image registration are the Moravec [8] and Harris [2] corner detectors. But, they are not the scale invariant registration methods that can avoid camera zooming. Mikolajczyk [6] refined Harris corner detector that detects interest point first, and computes Gaussian derivatives on each interest point, so that it can compare the feature points in different scale. Later, D.G. Lowe [5] presented SIFT (Scale-Invariant Feature Transform) algorism using Difference of Gaussian (DoG) to search feature points in different scale space. But the computation of this method is too complicate that the searching speed is very slow. Y. Ke and R. Sukthankar [3] used PCA (Principle Component Analysis) and designed a PCA-SIFT method to reduce the feature space dimension and speed up more.

Bay et al [1] proposed an interest point detector-descriptor scheme, called SURF (Speeded-Up Robust Features). The detector is based on the Hessian matrix, but uses a Laplacian based DoG (Difference of Gaussian) approximation. It relies on integral images to reduce the computation time. This way is faster and more robust than SIFT algorism. SURF is a scale and in-plane rotation invariant detector and descriptor. We use this algorism in this paper for image matching between sample and target banner ads.

III. SYSTEM ARCHITECTURE

For banner ads matching, interest point features are first extracted from the sample ad images and stored in a database. A new video frame is matched by individually comparing each point feature from the new frame to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors. In order to achieve this task, our system is divided into two execution phases, as shown in Fig.1, the Feature Learning Phase and the Banner Searching Phase.

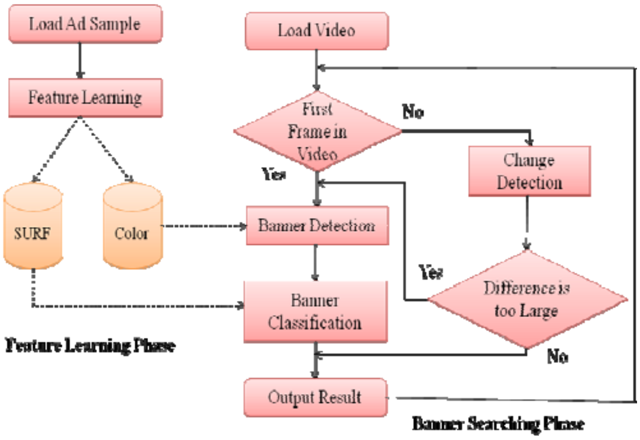


Fig.1 System architecture

In the Feature Learning Phase, the first thing needs to do is to choose the specific sample ad images. The second step is to collect both color histogram and SURF features of those sample images for the later matching process.

In the Banner Searching Phase, after loading the target video, the major searching and counting processes will be the Banner Detection and the Banner Classification.

IV. FEATURE LEARNING

In the Feature Learning Phase, as shown in Fig.2, our system needs to extract the color and SURF features from those samples ad images. After loading sample images into system, next step is to create the hue color histograms from HSV color space as the color features of those sample ads. In the other hand, our system uses the same procedures to extract the SURF features [1][4] from those samples.

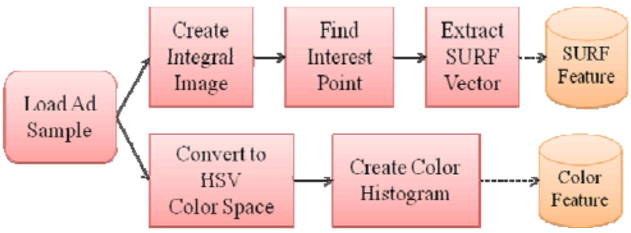


Fig.2 Flowchart of Feature Learning Phase

These SURF features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. They are well localized in both the spatial and frequency domains, reducing the probability of disruption by occlusion, clutter, or noise. The major stages of computation used to generate the set of sample ads features are (1) scale-space extrema detection, (2) interest point localization, (3) domain orientation assignment, and (4) interest point descriptor. Totally 68 dimensions of feature vectors are used for feature computation and matching, including the xy coordinate, the orientation, as shown in Fig. 3, the sign of Laplacian and the 64 features within the interest point neighborhood from the SURF descriptor.

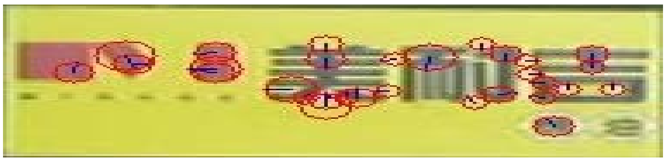
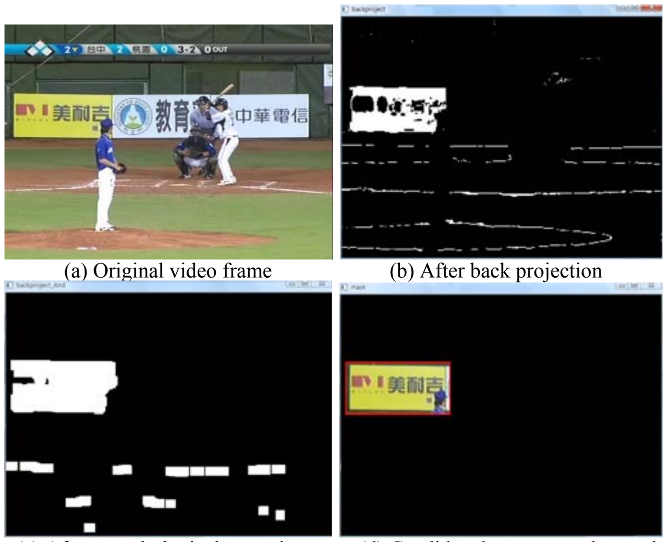


Fig.3 The position and orientation of detected interest feature points

V. BANNER DETECTION

In order to speed up the searching process, we first check the video frame weather the color feature of those sample ads exit or not. We change input video frame by the back projection function according to the color feature histogram. This back projection function puts the value of the histogram bin, corresponding to the number of color in the original video frame. In other words, the value of each video frame pixel is the probability of the selected number of color given the distribution (histogram).

After the bi-value thresholding and morphological operations, we can get rid of most of the noise and small area with the same feature colors as those sample ads from the video frame. The remaining parts of areas in this detection process, as shown in Fig.4 (d), are the places of candidate banner ads (also called ROI, Region of Interesting).



(a) Original video frame (b) After back projection
(c) After morphological operation (d) Candidate banner area detected

Fig.4 The results of banner detection

VI. BANNER CLASSIFICATION

A flowchart for banner classification is shown in Fig.5. Different scale of banners and same color in the scene will affect the recognition. For improving accuracy and efficiency, each detected ROI needs to be normalized [10] first. This normalization process can adjust the ROI to have an equal size with the sample ads.

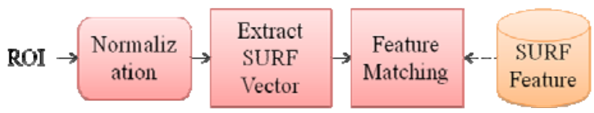


Fig.5 Flowchart of banner classification

After the extracting SURF feature vectors form the selected interest points in ROI, our method will go to the matching process. The correct matches can be filtered from the full set of matches by identifying subsets of interest points that agree on the object and its location and orientation in the ROI. TABLE 1 is the SURF feature points.

TABLE I. SURF feature points

detector						
151.000	x-location of the interest point					
23.000	y-location of the interest point					
69.6˚	orietation of the interest point					
-1	sign of Laplacian					
descriptor						
1~10	11~20	21~30	31~40	41~50	51~60	61~64
-0.003101	0.002867	0.418820	0.009239	0.027483	0.010100	0.000390
-0.002853	0.000823	-0.017696	0.004079	-0.016533	0.003207	-0.000137
0.008632	-0.004040	0.454265	0.026780	0.052236	-0.007098	0.000476
0.007705	0.001356	0.301441	0.020167	0.027360	-0.006072	0.000315
0.019864	0.004981	0.010591	0.046384	0.001277	0.085254	
-0.005101	0.001857	0.005243	0.058395	-0.000606	0.030851	
0.027061	-0.038449	0.015678	0.235045	0.001925	0.044847	
0.013249	-0.021729	0.015487	-0.070703	0.001500	-0.024600	
0.002698	0.052255	-0.005552	0.435966	-0.000557	0.047717	
-0.000706	0.058587	0.003656	0.490900	0.002701	0.028479	

System compared each sign of Laplacian that is to decide the value of Hessian matrix as in (1) is the maximum/minimum or not. Fig.6 presents the sign of Laplacian matching. It can save many time and calculation before matching SURF features.

$$Trace(H_{approx}) = D_{xx} + D_{yy} \tag{1}$$

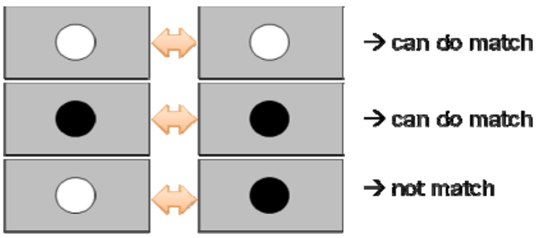


Fig.6 Sign of Laplacian matching

In feature matching of SURF, it can use many different functions of distance to measure the similarity between two interest points. The best candidate match for each interest point is found by identifying its nearest neighbor in the database of interest points from learning images. The nearest neighbor is defined as the interest point with minimum Euclidean distance from the descriptor vector.

A good interest point usually differs from neighbors greatly. So system will find the Euclidean distance between interest points of video frame and sample ads, and using NNDR (Nearest Neighbor Distance Ratio) [7] for interest point matching. Unfortunately, there always exists some interest points that have higher similarity to the real one, cause the matching point position not correct.

As shown in Fig.7, the sample ad v is in the upper left corner, and the detected candidate banner ads region w of video frame is in the lower right corner. The v1~vn are the interest points of sample ad v, and the w1~wm are interest points of in the ROI *w*. Each line in this figure represents a best matched pair of interest point in these two images.

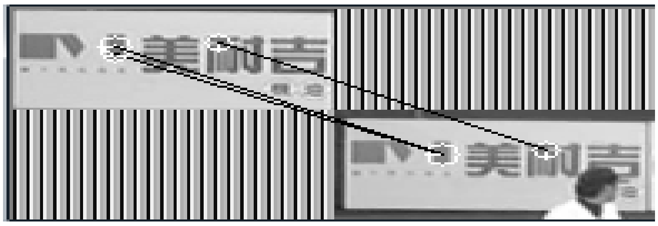


Fig.7 The corresponding points between sample ad and ROI

Therefore, the best matched pair *(v_i, w_j)* can be calculated by the Euclidean distance defined as:

$$dist(v_i, w_j) = \sqrt{\sum_k (\mathrm{v}_{ik} - \mathrm{w}_{jk})^2}\,, \tag{2}$$

Where vik is the k-th descriptor vector of the i-th interest point of sample ad v, and wjk is the k-th descriptor vector of the j-th interest point of ROI w. That is, dist(vik, wjk) is to calculate the Euclidean distance between the corresponding 64 descriptor vectors of the interest points.

Next, we can find a point p1st, i is the smallest distance in ROI w to point i in v. Also, we can find a point p2nd, i is the second nearest distance in ROI w to point i in v.

$$p_{1st,i} = \arg\min_{j} \left(dist(v_i, w_j) \right)\,, \tag{3}$$

$$p_{2nd,i} = \arg\min_{j,\, j \neq d_{1st,i}} \left(dist(v_i, w_j) \right)\,, \tag{4}$$

Then, we can accept v_i and w_j as a best matched pair if :

$$d_{1st} / d_{2nd} \leq \alpha \quad \text{And} \quad |\theta_1 - \theta_2| \leq \beta\,, \tag{5}$$

Where d1st, and d2nd, are the distance between vi and the first point wj and also the second point, α is a threshold value to ensure these two candidate points in the ROI w are separated far enough. θ1 and θ2 are the orientation direction of vi and wj, β is a threshold value to ensure these two points vi and wj have almost the same direction.

Because the interest point descriptors are highly distinctive, it allows a single feature to find its correct match. So that, a small amount of interest point pairs have been accepted, we can say a specific banner ad is displayed in this video frame and increase the counter of this ad by one.

Before we start to calculate the similarity between each pair of interest points in sample ads and ROIs, we first compare the sign of Laplacian of each pair that can decide whether these two interest point are the same kind or not. We can see in the Fig. 8 and Fig.9 presents the difference of

matching when using the sign of Laplacian. As shown in Fig.9, those large numbers of parallel line can show the result of a correct match.

Fig.8 The result of matching without using the sign of Laplacian

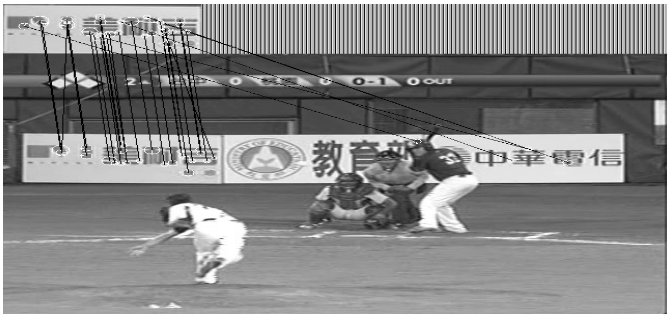


Fig.9 The result of matching using the sign of Laplacian

In order to speed up the searching process, when we check the next frame, we only compare the difference in the ROI areas between previous checked frame and current frame. If the difference is small, that means no significant change happened between these two frames, we don't have to do the matching process again. We can simply inherit the result of the previous frame, and go to the next one.

VII. EXPERIMENTAL RESULT

We design a user interface, as shown in Fig.10, for this banner searching system. The left middle area is the main playback region. We can open a video file and review in this area. The upper right area is the sample ad images region. We can select those specific sample ads from image files. The middle right areas provide the information about the total frame number of each sample ad matched, the time code of each matched ad, and the similarity value (distance) when execute the matching process.

The sports videos we select for this experiment are base on baseball and volleyball games. We analyze the accuracy by precision rate and recall rate, and also time spending for searching those sample banner ads in those videos. As shown in TABLE 2, the precision rates of baseball games are almost 100% and the recall rate are exact 100%. But in the volleyball games, because both the camera and players are moving too fast, it causes the motion blur in video and of course makes some mismatching in the results. We also can see form TABLE 2 that our system can search 27.1 frames per second in average. We can say it almost fulfills the real-time requirement.

Fig.10 The system interface

TABLE II. EXPERIMENT RESULTS

No	Category	Total frames	TP	FP	FN	Precision Rate	Recall Rate	Time(s)
1	Baseball	1800	1348	0	0	100%	100%	69.44
2	Baseball	1800	567	0	0	100%	100%	61.11
3	Baseball	1800	846	4	0	99.53%	100%	62.75
4	Baseball	1800	1210	4	0	99.67%	100%	63.74
5	Baseball	1800	610	0	0	100%	100%	67.47
6	Baseball	735	591	0	0	100%	100%	24.94
7	Baseball	1152	721	0	0	100%	100%	40.23
8	Volley ball	1800	1113	59	55	95%	95.3%	72.45
9	Volley ball	1800	634	65	64	90.7%	90.8%	72.84
Average		**53.66 sec.**	**7560**	**211**	**119**	**98.32%**	**98.45%**	**59.44**

*TP: True Positive, FP: False Positive, FN: False Negative.

As shown in Fig.11, and Fig.12, our system still can work correctly even if the banner is partially occluded by players, and also in the case of some other banners have the same dominate color distribution. In Fig. 11, we can see the banner in volleyball game scene is much smaller than baseball and can be recognize well.

Fig.11 The banner partially occluded by the players

Fig.12 Background and banner have the same color.

Fig.13 The banner is occluded by the subtitle and oblique

In Fig .13 and Fig.14, when the camera is moving fast and the scene is changing, the system will make some mistakes.

Fig.14 The scene is blurred

Fig.15 The scene is dissolved

VIII. CONCLUTIONS

This paper provides an approach to design a system for searching and counting the number of banner ads appearing in sports videos. The searching proceedings of matching feature points from pre-specified sample ads use the colour histogram and the SURF algorithm feature vectors. We define an acceptance criterion to select the best matched pair of interest points. We also use the frame difference to check a scene change to reduce computation time and to improve system speed. This approach for searching banner ads can robustly identify targets among scaling and occlusion while achieving nearly real-time performance. This system works effectively and has a very high correct rate in our experiments using real videos.

In the future, it also can be used in content-based image retrieval. Then, we could find the most unique feature of each image by learning any features of images and removing the same features. It can be much faster and improve the correct rate.

REFERENCES

[1] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," *Proc. of European Conf. on Computer Vision*, pp. 404-417, 2006.

[2] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. of Alvey Vision Conf., Manchester* , pp. 147-151, 1988.

[3] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 511-517, 2004.

[4] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," *IEEE ICIP*, vol. 1, pp. I-900 – I-903, 2002.

[5] D. G. Lowe, "Object recognition from local scale-invariant features," *Intl. Conf. on Computer Vision,* vol. 60, pp. 1150-1157, 2004.

[6] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *Intl. Conf. on Computer Vision*, pp. 525-531, 2001.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 27, no. 10, pp. 1615–1630, 2005.

[8] H. P. Moravec, "Towards automatic visual obstacle avoidance," *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, p.584, 1977.

[9] Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp.1349-1380, Dec 2000.

[10] S. J. Wang and Y. S. Wu, "Color image retrieval using a fuzzy dominant color selection clustering algorithm," *Journal of C.C.I.T*, vol.37, no.1, Nov. 2008.